HOMEWORK #1. PSY-GS 8888/ 3732 LATENT GROWTH CURVE MODELING FALL '15 STERBA

Total points=100

DUE IN CLASS SEPT. 23RD

Please answer the following questions. Note: I have given you the text dataset you will need (read_bywave.dat).

You are interested in the development of reading skills over time. You have access to a dataset containing N=221 children whose reading recognition was assessed with the Peabody (!) Individual Achievement Test across T= 4 waves of data collection (1986, 1988, 1990, 1992) in the National Longitudinal Study of Youth. Children were between 6-8 yrs. in 1986. You wish to consider the following models:

MODEL A: unconditional linear LGM with fixed intercept; fixed slope; equal level-1 error variances across time

MODEL B: unconditional linear LGM with random intercept; fixed slope; equal level-1 error variances across time

MODEL C: unconditional linear LGM with random intercept; random slope; correlated intercepts and slopes; equal level-1 error variances across time

MODEL D: unconditional linear LGM with random intercept; random slope; correlated intercepts and slopes; unequal level-1 error variances across time

(1 --13pts) Please write out the matrix equation for Models A-D, identifying the dimensions of each matrix and identifying the elements within each matrix. Equations can be hand written or typed, but the other responses in the writeup need to be typed.

(2 --8pts) Please fit Models A-D in Mplus (Demo version of Mplus should suffice for this assignment). Please include your input and your output for each model in a separately stapled packet at the end of your writeup and please write on the top of each of these pages YOUR LAST NAME, MODEL XX. Please clip these model inputs/output packets to your writeup.

(3 --22pts) For each model, interpret all estimated parameters, and list the number of *knowns* and *unknowns,* as discussed in class. Your interpretation of parameter estimates should be appropriate for a journal article results section. By this I mean that (a) you should provide a brief definition ($\leq$ 1 sentence) of each parameter, in words, the first time it is discussed in the writeup, and (b) when you report the numerical value of a parameter estimate in the text you should also explain what that numerical value means in the context of the substantive topic at hand.

(4 -- 22pts) For each model, interpret *overall* model fit using some or all of the relevant indices discussed in class for evaluating a model in isolation. Your interpretation of model fit indices should be appropriate for a journal article results section. By this I mean that (a) you should provide a brief ($\leq$ 1 sentence) definition, in words, of what the fit index evaluates the first time it is discussed in the writeup, and (b) when you report the numerical value of the fit index in the text, you should also explain what its implications are for the substantive topic at hand.

- Suppose our *N* was 2,221 rather than 221. All else equal, what would happen to the model chi-square statistic? Why is it important to consider some fit indices that are less sensitive to *N* and what is an example of such a fit index?

(5 --22pts) *Compare* the overall fit (not component fit) of Models A vs. B, B vs. C, C vs. D using some or all of the model selection procedures discussed in class.

- Clarify whether a given pair of models is *nested*; if nested, clarify which is the *less restricted model* and which is the *more restricted model* of the pair. Mention whether the model selection indices you are using require that models under consideration be nested.
- Interpret numerical values from model comparisons in words and state their implications for the substantive topic at hand. Briefly state your final conclusions regarding which model you think is most likely to have generated the data.

(6 --13pts) To more thoroughly evaluate component fit, for Model D only, inspect and report what was referred to as the *residual mean vector* and *residual covariance matrix* in the Component Fit section of Lecture 4. Note: you may choose to use either raw, standardized, or normalized residuals, but report which you used. In the case of raw residuals, we discussed that the former vector tells us the difference between observed and model-implied means $\left(\bar{\mathbf{y}} - \boldsymbol{\mu}_y(\hat{\boldsymbol{\theta}})\right)$ and the latter matrix tells us the difference between observed and model-implied (co)variances $\left(\mathbf{S} - \boldsymbol{\Sigma}_y(\hat{\boldsymbol{\theta}})\right)$. (Recall that the latter matrix is *much* different from $\boldsymbol{\Theta}_\varepsilon$, although both may unfortunately at times be called residual covariance matrices, for short.)

- Provide a brief (e.g., 1 sentence) possible interpretation of your findings from visual inspection of this vector and matrix. You may also wish to plot $\bar{\mathbf{y}}$ vs. $\boldsymbol{\mu}_y(\hat{\boldsymbol{\theta}})$ to aid interpretation, as discussed in Lecture 4.
- Please mention what you would infer about your model if all elements of the residual covariance matrix were 0, making reference to the concept of a discrepancy function.